STATDISK User Manual
Written for STAT 3001 (Statistical Methods)

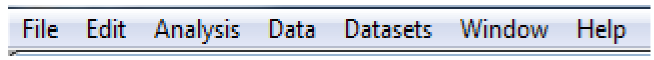
College of Undergraduate Studies

First Version written by Dr. James A. Condor (2011)
Updated by: Anne Keller Geraci (Feb. 2013)

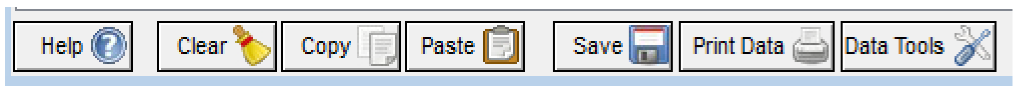
Contents

Basics of STATDISK

You can perform all STATDISK functions from the Sample Editor Screen using the following menus: **File, Edit, Analysis, Data, Datasets, Window**, and **Help**.



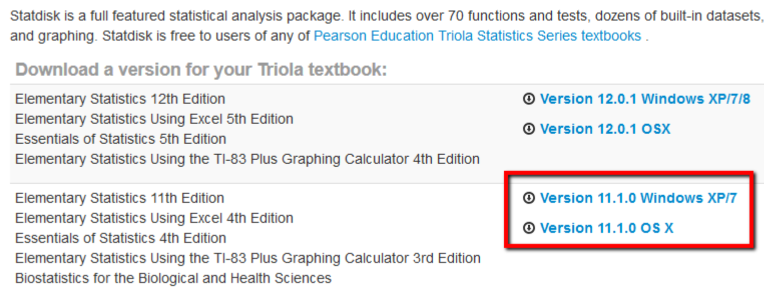
Along with performing statistical calculations, STATDISK is also compatible with many popular application software packages. You can import, copy, paste, save, print and transform data sets. You can also copy, paste, save, or print any of the STATDISK numerical or graphical outputs and export them into other programs such as Microsoft Word. Those options are available as clickable buttons at the bottom of the **Sample Editor** screen.



Downloading and Installing STATDISK

Use your browser and go to the website www.statdisk.org. Download Version 11.1 for your computer (Windows or OSX).

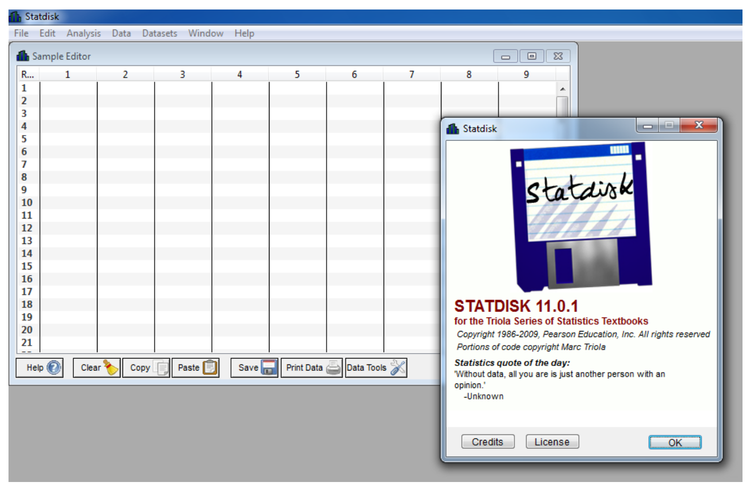
Once the file is downloaded, you will need to EXTRACT the files contained in the install package. On most computers, this requires a right-click with the mouse, then select Extract All and provide a location for the extracted files.



Once the install package is unzipped, you will need to find the application program **Statdisk**. Unlike other programs that need to be installed on your computer, STATDISK is just a file.

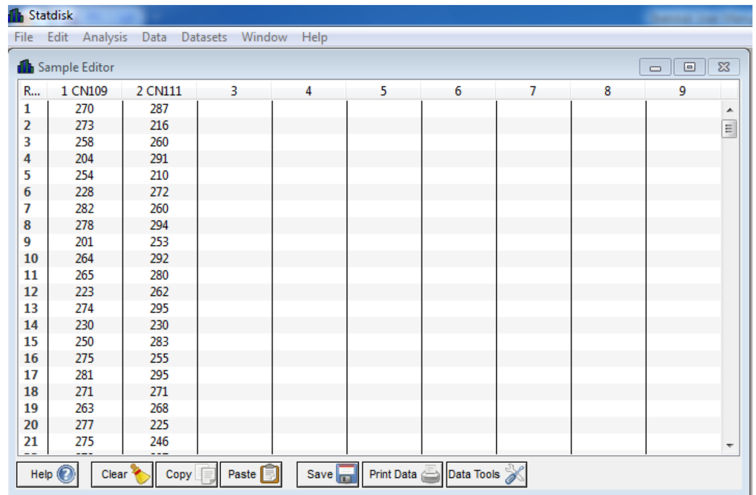
You might want to create shortcut for this file and place it on your desktop by right clicking with the mouse and selecting "Create Shortcut"

When you open the STATDISK program (by a double-click with your mouse on the above file) you will see the screen shown here. Click on the OK button to close the STATDISK information screen.



Opening a Data File

STATDISK has numerous datasets stored in the program and can be accessed by clicking on **Datasets** at the top of the **Sample Editor** window. After opening **Datasets** go to **Elementary Stats 9th Edition.** The names of the datasets will appear to the right. Click on **Cans** and the data values will appear in the **Sample Editor** as shown below.



You can preview the datasets before you open them by going to **Datasets** and then **Dataset Browser.** You can also access datasets that STATDISK has available online by going to **Datasets** and then **Online Datasets.**

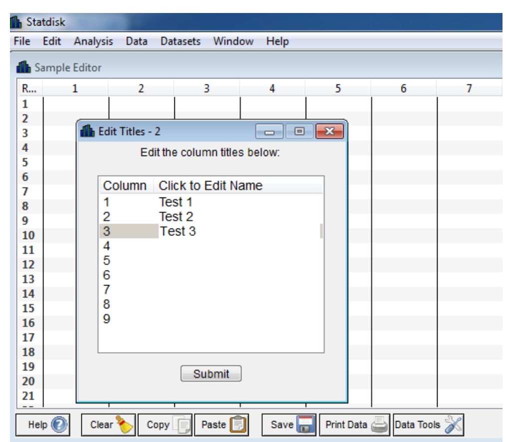
Using Data Tools

After you have opened a dataset or have typed in data to the **Sample Editor**, you can edit column titles, sort data, delete columns, add columns or rows, or explore the data set by opening the **Data Tools** menu.

The **Data Tools** button is located at the bottom of the **Sample Editor** page.

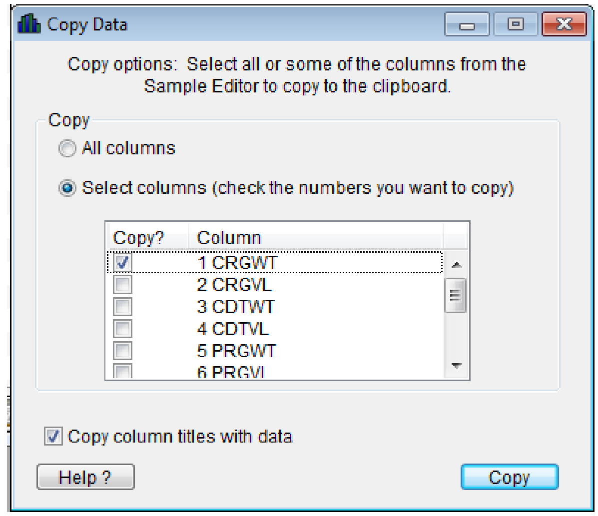
To Edit column titles open up **Data Tools** and then **Edit column titles.** Type in the names of the column titles into the box shown to the right.

Click on the **Submit** button to enter the new column titles.



Copy and Paste

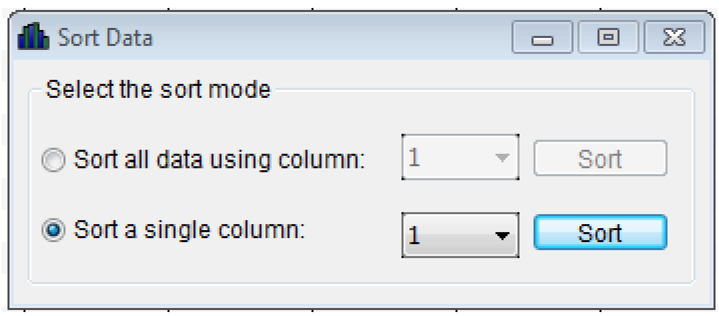
The Copy and Paste buttons are on the bottom of the **Sample Editor** Screen.



To copy or paste a data set simply click on the desired button and a screen will appear asking you which column of data you are working with. You can copy all of the columns or select columns. The **Paste** button directions are the same as the **Copy** button directions.

Sort Data

To sort data, open the data tools and select **Sort data.** Select **Sort a single column:** and then use the drop-down arrow to select the column of data values that you want to sort Then click on **Sort**. The data values in that column will be sorted from lowest value to highest value.



Saving your data

Save your data by clicking the SAVE button at the bottom of the Sample Editor screen. Provide a filename for the file, as well as the location on your computer where it should be saved.

The Data Menu

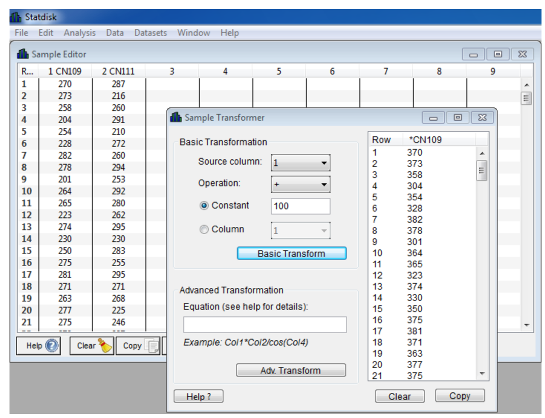
The two menus in STATDISK that are used to perform statistical procedures are **Analysis** and **Data.**

The **Data** menu is used to bring up the **Sample Editor**, transform data, sort data, generate descriptive statistics including charts and graphs, assess normality and generate sets of data values that emulate one of the standard types of statistical distributions.

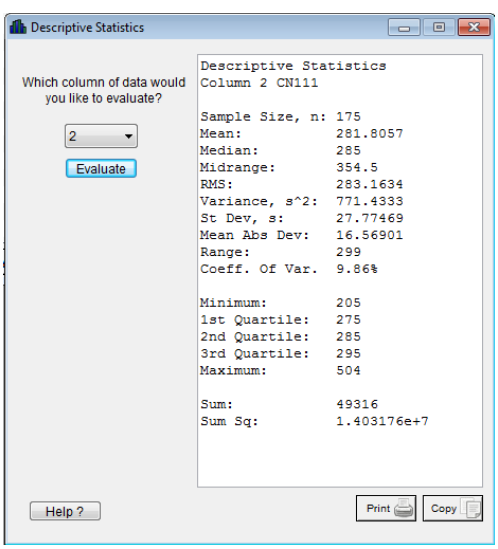
The **Analysis** menu is use to find area under the curve for many of the standard statistical distributions, determine sample size, create confidence intervals, perform hypothesis tests for parametric and non-parametric models.

Using the Data Menu

To transform a dataset you first need to type data into the sample editor or select an existing dataset. Open the **Cans** dataset. Select **Data** and then **Sample Transformations** to open the **Sample Transformer** window. The Source column is the column containing the dataset that you want to transform. Select the operation that will be used to change the data values and type in the constant that you will add, subtract, multiply, divide, mod value, or raise to a power to the data values. After you click on **Basic Transform** the new data set will appear in the **Sample Transformer** window. Now use Copy/Paste to transfer your transformed data into your editor.

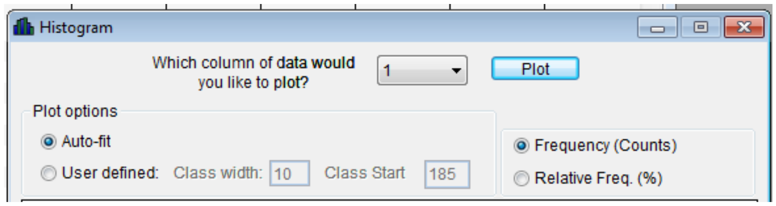


Descriptive statistics for a data set can be computed by opening the **Data** menu and selecting **Descriptive Statistics.** Select the column that the data set is in and then click on **Evaluate.** A list of the most commonly used numerical descriptive statistics will be displayed, as shown.



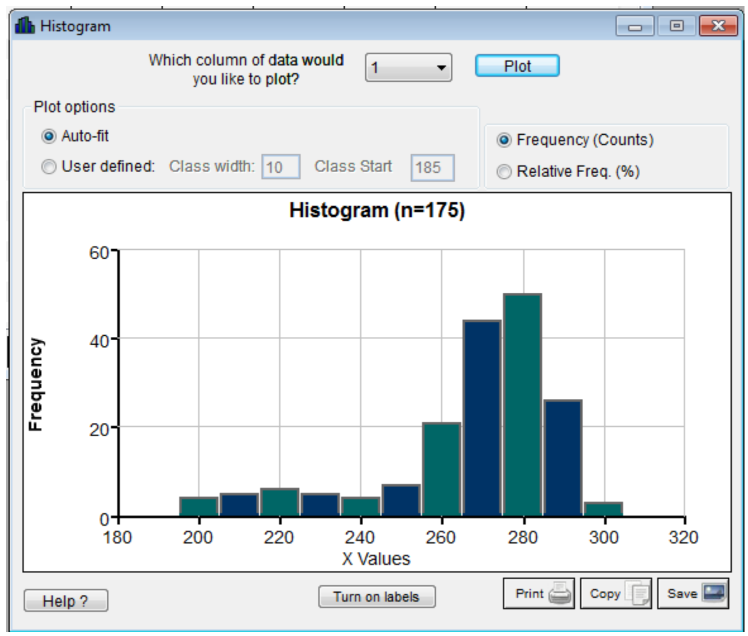
Histogram

A visual display of a single set of data values can be shown by opening the **Data** menu and then selecting **Histogram**.



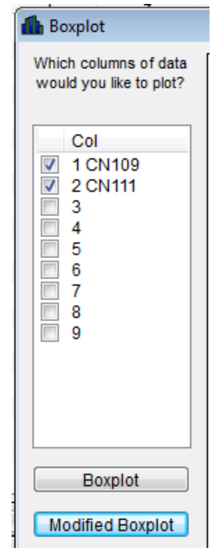
Select the column that the data values are in. If you would like the STATDISK program to automatically select the class width and the class start, select Auto-fit. You can display the count or the frequency for each class. Click on **Plot** to display the graph.

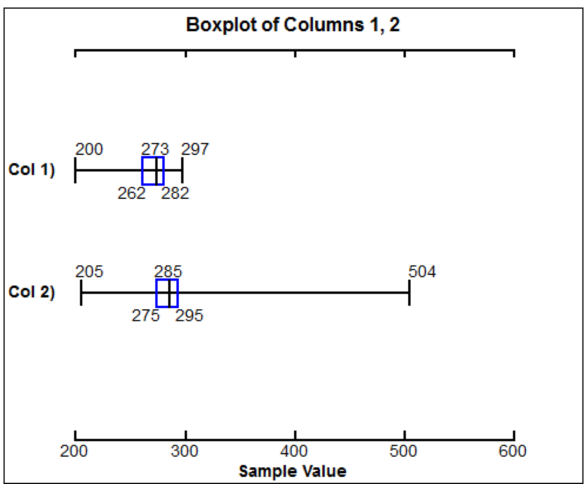
To display the counts or frequencies for each bar, click on the **Turn on labels** button at the bottom of the screen.

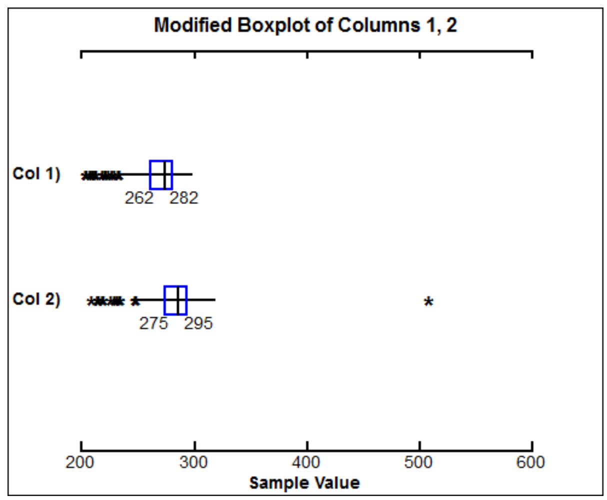


Boxplots

If you would like to compare two or more sets of data values you can plot them on one graph by using boxplots. Open the Data menu and select **Boxplot**. Then select the columns containing the data values that you would like to compare. You can then select **Boxplot** to show a standard view of the boxplots or **Modified Boxplot** which will emphasize outliers (see figure 11).







Basic Statistical Functions

STATDISK can perform many basic statistical functions relating to probability distributions, confidence intervals, hypothesis testing, correlation and regression, Chi-square and other non-parametric tests, and sample-size determination. This section will explain how to perform many of those basic statistical functions.

Normal Distribution

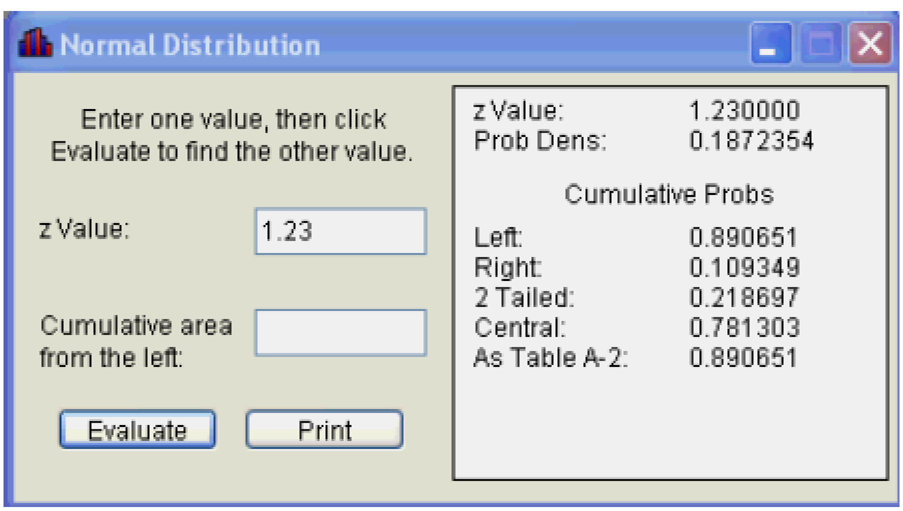
STATDISK uses standard z scores, so first convert scores by using

$$z = \frac{\text{data value} - \text{mean}}{\text{standard deviation}}$$

Here is the STATDISK procedure for finding areas or values from a normal distribution.

1. Select **Analysis** from the main menu at the top of the screen.
2. Select **Probability Distributions** from the subdirectory.
3. Select Normal Distributions.
4. Either enter a standard z score or enter the known cumulative area to the left of a z score.
5. Click on Evaluate.

For example, if you enter a z score of 1.23 in Step 4 above, the STATDISK display will be as shown below.



This display shows that the area to the left of $z = 1.23$ is 0.890651, and the area to the right of $z = 1.23$ is 0.109349. You may ignore the reference to Table A-2, because that reference applies to books in the Triola Statistics Series.

The chart below shows the standard normal distribution with Z-values along the bottom axis and the area under the curve between the given Z-values and can be used for full or half increments of the standard deviation. STATDISK will find the given values and any other values that are not shown on the table.

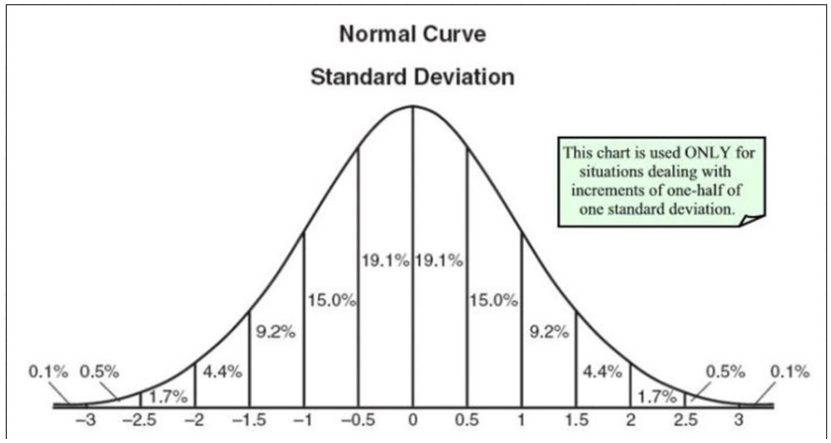


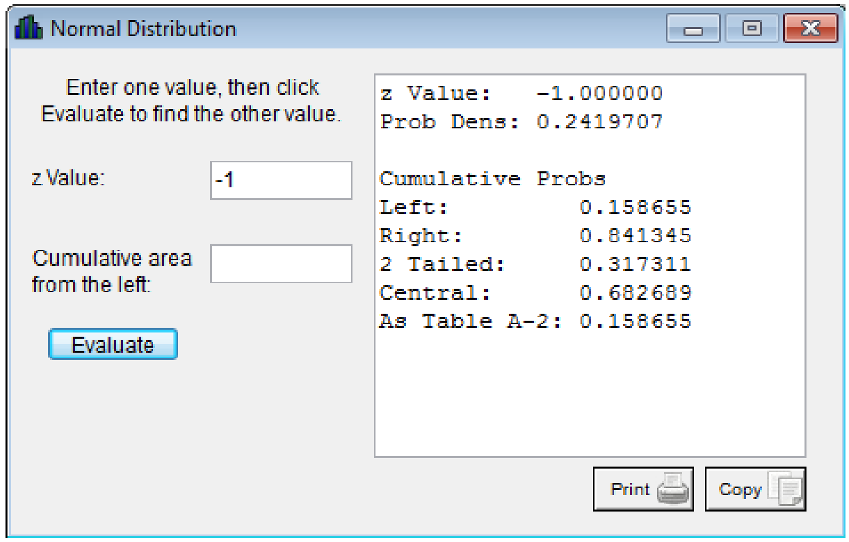
Figure 1. Standard Normal Distribution

Central Limit Theorem

Section 5.3 of *Statistical Reasoning for Everyday Life* discusses the Central Limit Theorem in detail. When using STATDISK, it is important to apply the Central Limit Theorem as follows:

When working with a sample of size *n*, compute the value of the standard *z* score by changing the standard deviation so that it is divided by the square root of *n*.

To find values that are not shown on the table, use STATDISK as follws:Open the **Analysis** menu and then select **Probability Distributions** and then **Normal Distribution**. Enter your z-score into the box for **Z Value** and then click on **Evaluate**. In this example, the z-score is -1 and the probability is 0.2419.

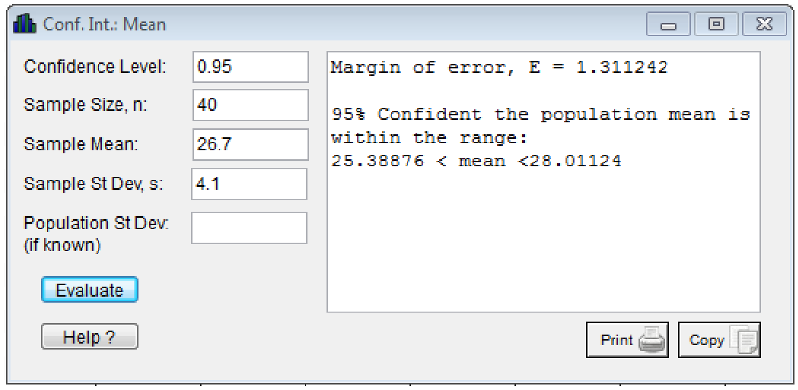


The output gives the discrete probability of seeing a z-score of -1 and is equal to .2419707. It also gives the cumulative area to the left of -1 or .158655. If you add the areas to the left of -1 shown in the standard normal distribution: 0.1% + -.5% + 1.7% + 4.4 % + 9.2% = 15.7% (or 0.157) you can see that you get the same result.

If you put in any value between 0 and 1 representing the area to the left of a Z score and then press **Evaluate** you will get the associated Z value.

Confidence Intervals

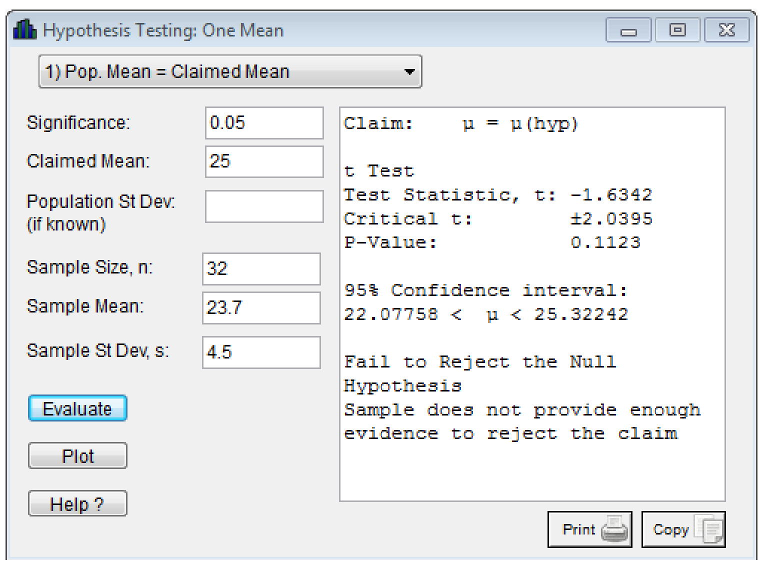
To find a confidence interval for a sample statistic you do not need to type in any data values or have a dataset in the **Sample Editor.** For example, to find a confidence interval for one-sample mean open up the **Analysis** menu then select **Confidence Intervals** and then **Mean-One Sample**. The image below shows the STATDISK output screen for a 95% confidence interval with a sample mean of 26.7, a sample standard deviation of 4.1, and a sample size of 40. The confidence interval of 25.29 to 28.01 is given. The Margin of error is the distance from the mean to the upper value and the distance from the mean to the lower value of the confidence interval.



If you are given a set of data values and not given any of the sample statistics such as the mean and standard deviation you must first use **Descriptive Statistics** to find the values needed to enter into the **Con. Int.: Mean** window that is shown in Figure 15.

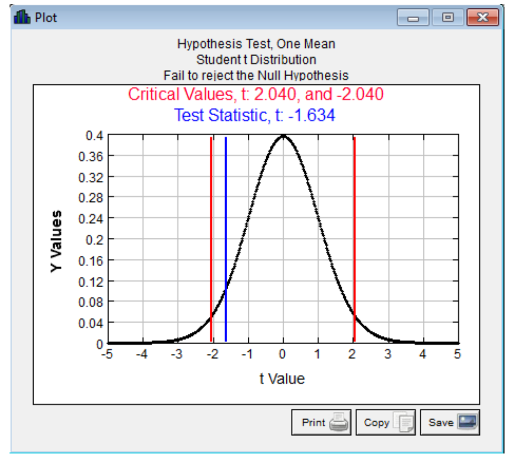
Hypothesis Testing – large sample

The hypothesis testing procedures in STATDISK are very similar to the confidence interval procedures. To perform a hypothesis test about a one-sample mean open up the **Analysis** menu and then select **Hypothesis Testing**, and then **Mean-One Sample**. Figure 16 shows the STATDISK output for a null hypothesis that the population mean is equal to the claimed mean, the hypothesized mean is equal to 25 and the sample mean is 23.7 with a sample standard deviation of 4.5 with a sample size of 32. The hypothesis is tested at the .05 level of significance. After you select **Evaluate**, you get the information shown. The information is provided on the right of the screen for the provided inputs.



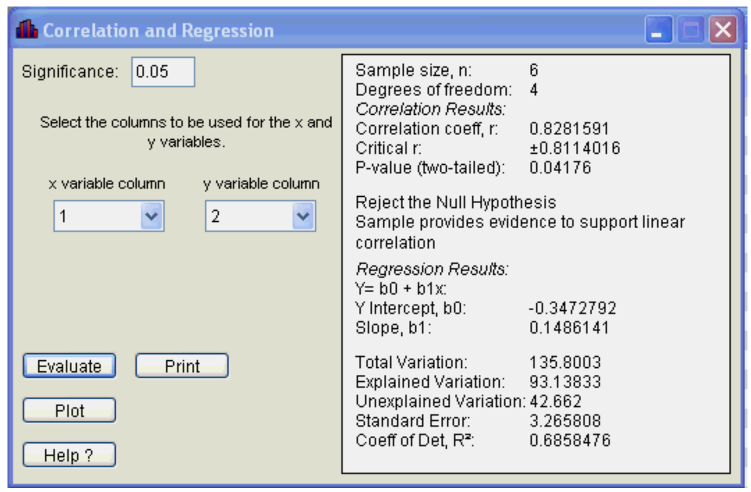
Note: As with confidence intervals if you are given a set of data values and not given any of the sample statistics such as the mean and standard deviation you must first use **Descriptive Statistics** to find the values needed such as sample mean and sample standard deviation.

To see a a normal probability plot for a given Hypothesis test, click the “PLOT” button. This will produce a graph that represents the visual interpretation of the hypothesis test, as shown here.



Correlation and Regression

To compute a correlation or create a regression equation you first need to type data into the **Sample Editor** or select an existing dataset. Open **Datasets** and select **Elementary Stats 9th Edition**. Open the **Homes** dataset.



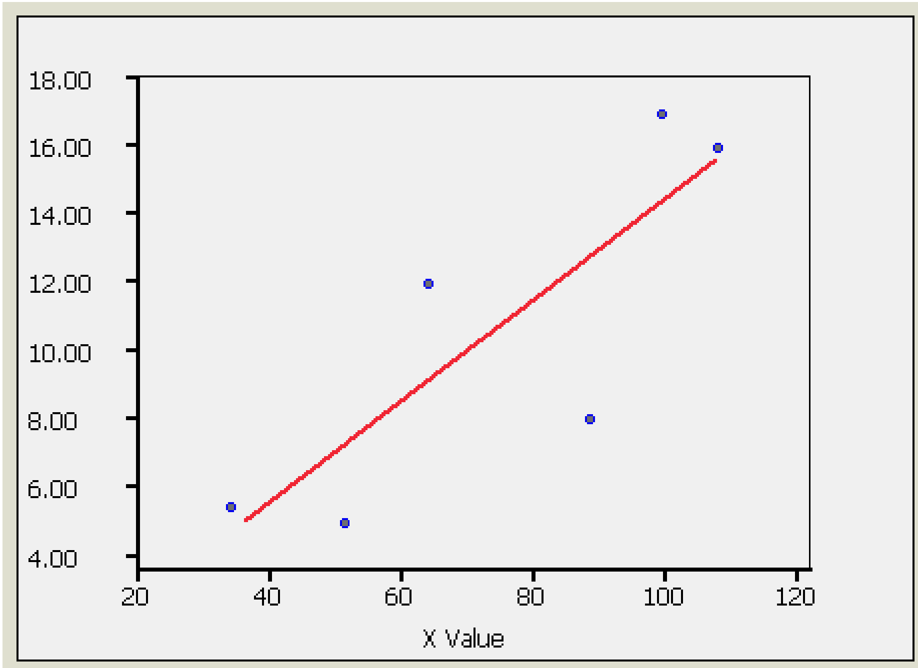
Select **Analysis** and then **Correlation and Regression.** Select the columns (2 and 3 in this example) to be used for the x and y-variables and then click on **Evaluate**. The information for both the correlation and the regression is shown in the output window.

Here are the key results from the above output:

- The correlation coefficient is $r = 0.8281591$.
- Based on the STATDISK result that "sample displays evidence that the variables are correlated," we see that the correlation is significant. (We could also refer to Table 7.3 in the textbook to find that the correlation is significant at the 0.05 level because the correlation coefficient of r = 0.8281591 is greater than the table value of 0.811. The correlation is not significant at the 0.01 level because the correlation coefficient does not exceed the Table 7.3 value of 0.917.)
- The y-intercept and slope of the best-fit line are −0.3472792 and 0.1486141, respectively. Based on these results, the equation of the best-fit line can be expressed (in the format of $y = mx + b$) as $y = 0.14861141x - 0.3472792$. (Largely for the reason of using a format that can be extended to include more variables, this equation is often expressed in this format:

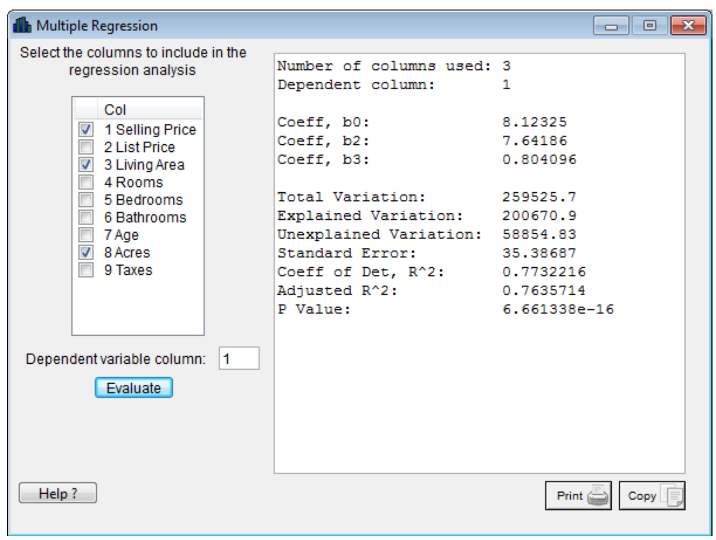
$$y = -0.3472792 + 0.1486141x$$

Shown below is the scatter diagram obtained when you click on "Plot" button. Note that the scatter diagram also includes the graph of the best-fit line. We get to see just how good the "best" fit actually is. In this case, there is a good fit because the data points are reasonably close to the best-fit line.



Multiple Regression

To generate a multiple regression equation you first need to type data into the **Sample Editor** or select an existing dataset. Open **Datasets** and select **Elementary Stats 9th Edition**. Open the **Homes** dataset. Select **Analysis** and then **Multiple Regression.** Select columns 1, 3, and 8 to be included in the regression analysis. Select 1 for the Dependent variable column. Click on **Evaluate** to generate the multiple regression statistics



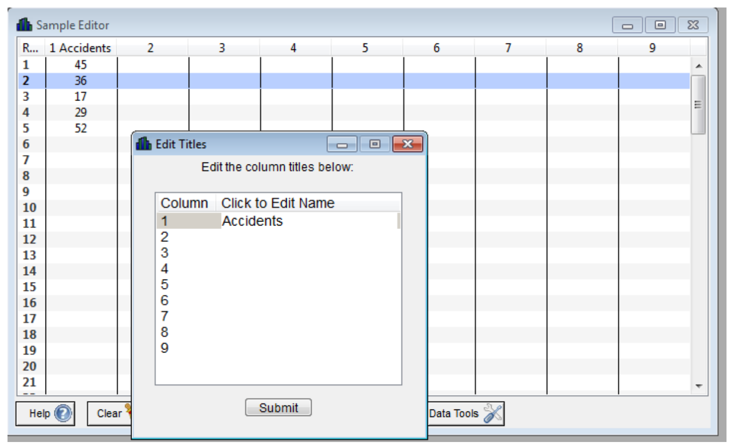
Additional Techniques

Chi-Square Goodness-of-Fit

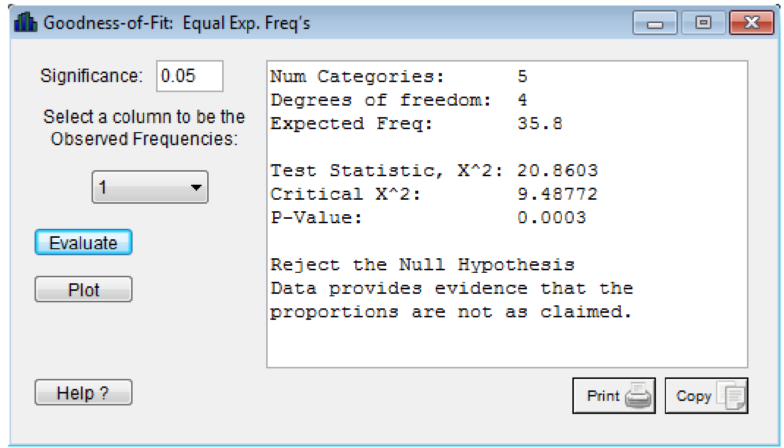
To generate a Goodness-of-Fit test to determine if you have equal expected frequencies, you must first type data into the **Sample Editor** or select an existing dataset. Let's imagine that a company wants to know if auto accidents occur equally throughout the days of the week. Use the **Clear** button at the bottom of the **Sample Editor** screen to erase any existing data. The number of accidents in our sample data that occur each day of the week are as follows:

M	T	W	TR	F
45	36	17	29	52

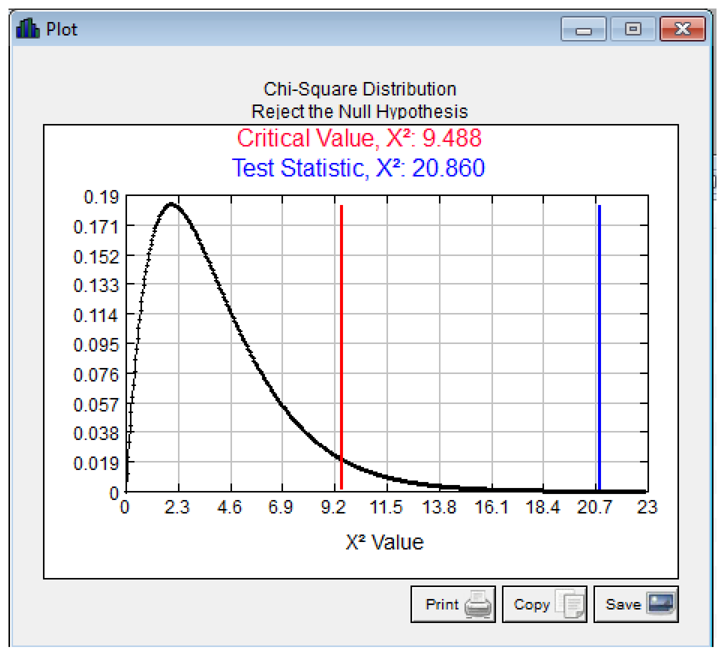
Type the data into List 1, then use the **Edit** column titles option under the **Data Tools** button at the bottom of the **Sample Editor** screen to name the variable: *Accidents*



Now select **Analysis** and then **Goodness-of-fit.** Chose **Equal Expected Frequencies** since the company is testing to see if accidents occur equally. Set the significance level to 0.05 and select 1 as the column to be the Observed Frequencies. Click on **Evaluate** to generate the Goodness-of-Fit test. The results are shown in the output window to the right.



Press **Plot** to view a visual representation of the Chi-Square Distribution of the data. The graph shows the Critical Value, X^2 : 9.488 and the Test Statistic, X^2: 20.860.



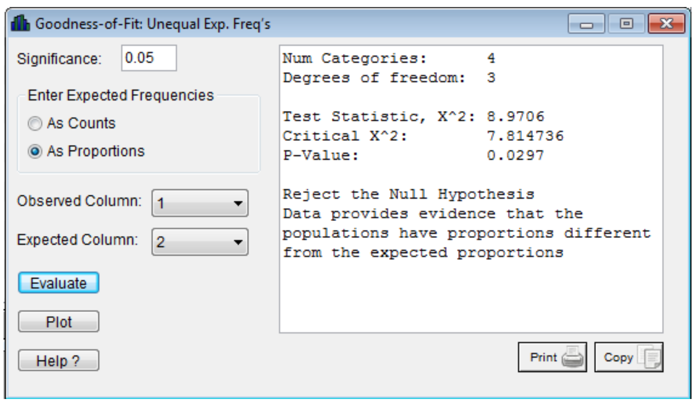
Goodness-of-Fit: Unequal Expected Frequencies

An ice cream company wishes to discover the popularity of their offered ice cream flavors. The Expected frequencies are given:

Vanilla	Chocolate	Strawberry	Other
42%	33%	14%	11%

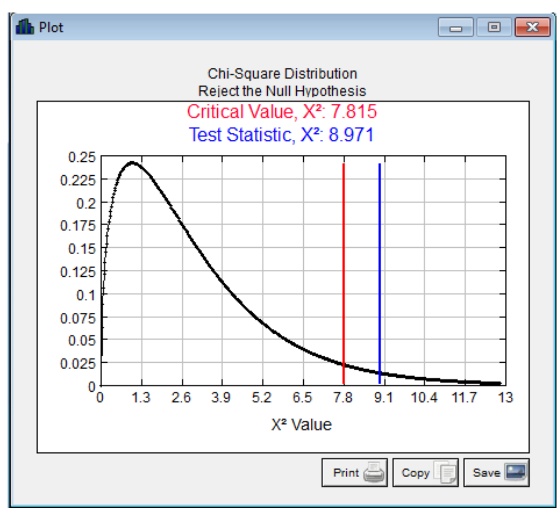
The University of Florida surveyed a sample size of n=250 students questioning their *preferred* ice cream flavor. The observed data collected is shown in the table below.

Vanilla	Chocolate	Strawberry	Other
114	68	47	21



In order to generate the goodness-of-fit test, the data must be entered into the Sample Editor. Use the **Clear** button at the bottem of the **Sample Editor** screen to erase any existing data. Enter the observed values into List 1 and enter the expected frequencies into List 2. Click on **Analysis** and then **Goodness-of-Fit**. Chose the **Unequal Expected Frequencies** option since the company is not testing to see if the flavors are equally popular. Because the expected frequencies were given as proportions, chose the **As Proportions** option under **Enter Expected Frequencies**. Set the **Observed Column** option as 1 and the Expected Column option as 2. We will set the Significance level to 0.05. Click **Evaluate**.

Click **Plot** to view a visual representation of the Chi-Square Distribution. The Critical Value, X^2 is shown as 7.815 and the Test Statistic, X^2 is shown as 8.971 (see Figure 26).

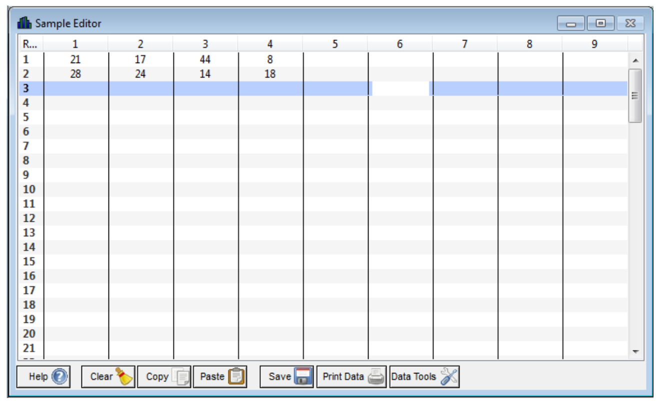


Chi-Square Test of Independence (Contingency Tables)

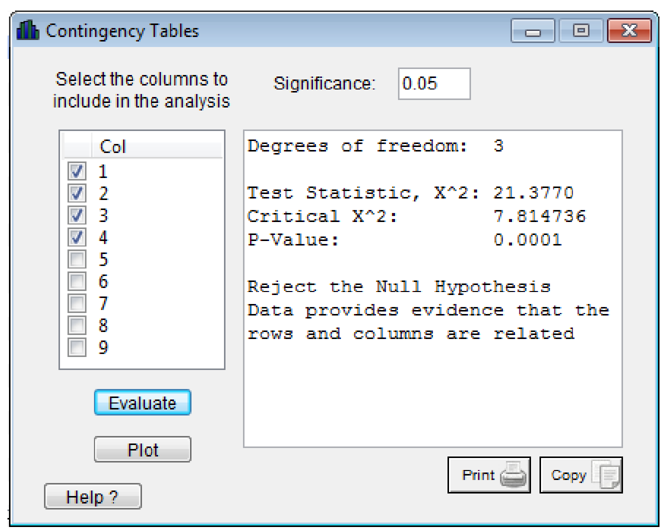
To generate a Contingency table test you must first type data into the Sample editor or select an existing data set. A company seeks to discover which color of car that males prefer and which color of car that females prefer. Use the **Clear** button at the bottem of the **Sample Editor** screen to erase any existing data. The data collected is as follows:

	Red	Blue	Green	White
Male	21	17	44	8
Female	28	24	14	18

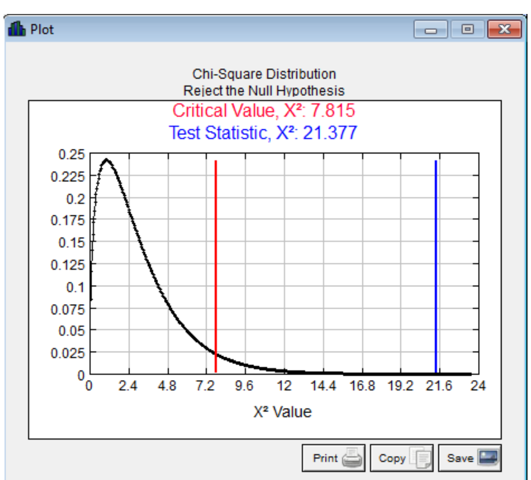
Enter the data into the Sample Editor exactly as it is shown in the table.



Select **Analysis** and then **Contingency Tables**. Then chose columns 1, 2, 3 and 4 to include in the analysis. We will set the significance level ot 0.05. Click **Evaluate** to view the results shown in the output window to the right (see Figure 28).



Click **Plot** to display a visual of the Chi-Square Distribution. The Critical Value, X^2 is shown to be 7.815 and the Test Statistic, X^2 is shown to be 21.377 (see Figure 29).



One-Way Analysis of Variance (ANOVA)

To use the Analysis of Variance (ANOVA) function in Statidisk you first need to type data into the sample editor or select an existing dataset. Open the **Homeruns** dataset. Go to the **Analysis** menu and then select **One-Way Analysis of Variance.** Select columns 1, 2, and 3 and click on **Evaluate**.

The hypothesis testing results are as shown.